CS60018: Statistical Learning Theory

End-Autumn Semester Exam

Time: 3:00hrs   Answer all FOUR questions.   Max marks: 100

1.A. Consider a hypothesis class $\mathcal{H}$ defined on the finite input domain $\mathcal{X}$. Let $k \leq |\mathcal{X}|$ be an integer. Every function $h \in \mathcal{H}$ outputs 1 for *exactly* k elements of $\mathcal{X}$. Find the VC dimension of $\mathcal{H}$. [10]

1.B. Define Rademacher complexity. What are the differences between Rademacher complexity and VC-dimension in characterizing richness of function classes. [10]

1.C. Define non-uniform learnability of a hypothesis class $\mathcal{H}$. If $\mathcal{H} = \{\bigcup \mathcal{H}_i\}$, what are the conditions such that $\mathcal{H}$ is non-uniformly learnable? [5]

1.D. Give an example of a function class defined on $\mathbb{R}$ which is non-uniformly learnable but not PAC learnable. [5]

2.A. Define weak-learnability of a hypothesis class $\mathcal{H}$. State the desired properties of a γ-weak learner L. Show that a class $\mathcal{H}$ is PAC learnable if it is weak learnable. [10]

2.B. Consider a two-class classification problem of predicting whether a student will get selected in a sports team based on his/her height h. The hypothesis class $\mathcal{H}$ consists of all hypothesis which predicts 'select' for a student if the height h is either less than a threshold β_1, or greater than another threshold β_2. In other cases the hypothesis predicts 'not select'. A simpler hypothesis class $\mathcal{H}_W$ uses a single threshold β, and predicts 'select' if h is greater than the threshold β, and 'not select' otherwise. Propose an ERM algorithm on $\mathcal{H}_W$ that is a γ-weak learner of $\mathcal{H}$. Obtain a value of γ. [10]

3.A. Define the Mistake bound and the Regret bound models of online learning. [10]

3.B. Define the Littlestone's dimension in online learning. [5]

3.C. Describe the HALVING algorithm for online learning. [5]

3.D. Consider the input domain of set of integers $\mathcal{X} = \{1, 2, \ldots, d\}$, where $d \geq 2$. Let $\mathcal{H} = \{h_j : j \in [d]\}$, where $h_j(x) = 1$, if $x = j$, and 0 otherwise. Derive a lower and an upper bound on the mistakes $M_{HALVING}(\mathcal{H})$ of the HALVING algorithm on $\mathcal{H}$. [10]

4.A. Define a convex function. [5]

4.B. State when a function is said to be ρ-Lipschitz. Show that the function $f(x) = \log(1 + \exp(x))$ is 1-Lipschitz over $\mathbb{R}$. [5]

4.C. Consider the problem of learning in logistic regression. The output probability $P(y = 1|x) = \text{sigmoid}(x) = \frac{1}{1+\exp(-wx)}$, is thresholded at 0.5 to obtain the output y. Regression weight parameter is w. The input domain is defined as $\mathcal{X} = \{x \in \mathbb{R} : ||x|| \leq B\}$ for some scalar $B > 0$. The label set is $\mathcal{Y} = \{\pm 1\}$ The loss function ℓ is defined as $\ell(w, x, y) = \log(1 + \exp(-y(w \cdot x)))$. Show that the resultant learning problem is convex and Lipschitz bounded. [10]